Calhoun
Institutional Archive of the Naval Postgraduate School

Calhoun: The NPS Institutional Archive

DSpace Repository

Faculty and Researchers | Faculty and Researchers' Publications

2015

Big Data Architecture and Analytics for Common Tactical Air Picture (CTAP)

Zhao, Ying; Kendall, Walter; Young, Bonnie; Germershauser, Zachary

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/57688

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun

DUDLEY KNOX LIBRARY

http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

NAVAL POSTGRADUATE SCHOOL

NAVAL RESEARCH PROGRAM

MONTEREY, CALIFORNIA

BIG DATA ARCHITECTURE AND ANALYTICS FOR COMMON TACTICAL AIR PICTURE (CTAP)

by

Ying Zhao, Ph.D. (PI), Information Sciences

Walter Kendall, Information Sciences

Bonnie Young, Systems Engineering

Zachary Germershausen (1st Lt), Student at GSBPP Program Office

10/1/2014 to 4/30/2015

Prepared for: Topic Sponsor OPNAV N2/N6 F33
Sponsor POC Jim Miller, CDR, USN & Mr. William Treadway
Integrated Fires Concepts
2000 Navy Pentagon, 1E275
Washington, DC 20350-2000
Office: (703) 695-8009
Cell: (904) 994-1278
NIPR: james.e.miller5@navy.mil
SIPR: james.e.miller@navy.smil.mil

FY15 MID-YEAR REPORT

Background

The amount of data generated by intelligence, surveillance, and reconnaissance (ISR) sensors has become an overwhelming challenge for decision makers involved in CID and the CTAP environment that Big Data architecture and Analytics (BDAA could address. The Navy needs to apply new architectures and analytics such as the current state-of-the-art BDAA). We will show how Big Data Architectures and Analytics (BADD) can collect and analyze the rising tide of sensor Information and fuse it in a timely manner to enhance Common Tactical Air Picture (CTAP).

For instance, more specifically, accurate, relevant and timely Combat Identification (CID) enables the warfighter to locate and identify critical airborne targets with high precision, optimizes the use of long-range weapons, aids in fratricide reduction, enhances battlefield situational awareness, and reduces exposure of U.S. Forces to enemy fire.

Specifically we want to study 1) how to identify and assess the current CID methods, for example, the best combination of platforms, sensors, networks, and data including organic sensors, regional sensors and National Technical Means (NTM) that can be used for track correlation and continuity of CID and correlate IDs to tracks regarding the state-of-the-art of the systems, applications, databases from Navy, Joint and National programs. 2) how to use massive parallel and distributed computation to improve CID's fidelity and latency reductions through recursive data fusion; 3) how to use learning agents to discover and learn the patterns in historical data and correlate the patterns with real-time data to detect anomaly; 4) how the discovered patterns might be consistent with all functional elements and existing rules for the Cooperative and Non-cooperative CID methods. 5) how to address the unique challenges of CID (e.g., extremely short dwell time for fusion, decision making, and targeting; uncertain or missing data outside sensor [e.g., radar, radio] ranges; and limited resources in air); 6) Improve real-time targeting recommendations and decision making towards a future vision of automated battleforce management.

Process

We are working on two key tasks as follows:

Task 1: Identify/assess the current CTAP, identify key elements required to support IAMD and NIFC-CA, for example, the best combination of platforms, sensors, networks, and data in a common tactical air picture.

Task 2: Investigate/explore how to apply innovatively the big data architectures and related analytic platforms such as Lexical Learning Agents (LLA) jointly with Hadoop, map/reduce, Distributed R, Rhadoop, Apache Mahout and RAD (Rapid Applilcation

Development) to such as Model View Controller (MVC) technologies to construct a better CTAP.

For Task 1,

- We reviewed a large collection of documents provided by domain experts and have periodic conference calls.
- We attended the National Fire Control Symposium (NFCS) and communicate with the CID domain experts to understand the current status.
- We are in the process of compiling a taxonomy for CID as a deliverable.

For Task 2

- We went to the Silicon Valley Big Data and Deep Learning meetups and seminars to understand the state-of-the-art technologies and methods in the area.
- We are also in the process to publish a white paper titled “Big Data and Deep Learning for Understanding DoD Data” to the Crosstalkonline.org Data Mining and Metrics, Jul/Aug 2015 Issue (http://www.crosstalkonline.org) which CTAP is presented as a use case.

Findings and Conclusions

We presented the initial results regarding how to apply BDAA analytics such as unsupervised learning, self-taught learning, deep learning, pattern recognition, anomaly detection and data fusion, with tools such as Lexical Learning Agents (LLA), System Self-Awareness (SSA) and Collaborative Learning Agents (CLA) jointly with Hadoop, Map/Reduce, Spark, Apache Mahout and traditional data sciences such as Model View Controller (MVC) to improving CID at the NFCS.

We found the potential and anticipated benefits by using BDAA to CTAP are listed as follows:

- Process more data sources in parallel
- Automate data fusion with heterogeneous databases to improve object recognition and predictive accuracy
    - For example, “Representation learning through topic models “((Dr. Arjuna Flenner, Dr. Arjuna Flenner, NAWCWD, China Lake) might be able to fuse heterogeneous data sources for CID
- Machine-learn associations for diversified attributes from different sensors which may not be in the same data formats for fusing heterogeneous databases
    - For example, Navy Tactical Cloud (NTC) requires a pre-defined meta-data, data standard, ontology and taxonomy to set up an application like Combat ID. Big Data technologies such as graph databases may be leveraged to facilitate analytics that might not correlate, associate and fuse heterogeneous data sources
- Perform pattern recognition and anomaly detection
    - Lexical Link Analysis (LLA), System Self-Awareness (SSA) and Collaborative Learning Agents (CLA)
    - Deep learning, machine vision, large-scale object identification across heterogeneous data sources

    - Mahout, Spark & Velox: Spark (Spark, 2015): Map/Reduce is an analytic programming paradigm for Big Data. It consists of two tasks: 1) the "Map" task, where an input dataset is converted into key/value pairs; and 2) the "Reduce" task, where outputs of the "Map" task are combined to a reduced key-value pairs. Apache Spark could replace Map/Reduce and Mahout for its speed and in-memory computation.
    - Bayesian Networks with R and Hadoop (Mendelevitch, 2015): it is a data-driven learning of conditional probability or structure learning. It is a supervised learning method but most good for Big Data with low dimensions. It is an approximate inference good for Big Data and Hadoop implementation.
    - Quora type of graph models might be useful too (http://data.quora.com)
- Enhance decision-making and resource management through automated data analytics
- Provide CID through custom models or through commercial software such as Model View Controller (MVC)

Recommendations

We recommend a next step to work with a couple of collaborators to identify a few data sources and a portfolio of DBAA methods to provide evidence, demo and use case for our findings. Specifically we will select test cases, data and beds to illustrate the advantages of BDAA to address the unique challenges of CID (e.g., extremely short dwell time for fusion, decision making, and targeting; uncertain or missing data outside sensor [e.g., radar, radio] ranges; and limited resources in air):

• Use massive parallel and distributed computation to improve CID's fidelity and latency reductions through a recursive data fusion. For example, to illustrate using the Map/Reduce-like paradigm to "map" or associate data sources that are not confined to standards (i.e. heterogeneous data sources) and "reduce" them to smart data that improve CID

• Use deep learning techniques for object and activity identification in massive images and videos

• Use learning agents to discover and learn the patterns in historical data and correlate the patterns with real-time data to detect anomaly

Approaches identified in this study will inform programs requiring BDAA support and inform POM 19. These issues have implications for N2/N6, N96, N98, NAVAIRSYSCOM, PEO(C4I), and PEO(IWS).